DOCUMENT RESUME

ED 220 528 TM 820 550

AUTHOR Faires, Charles L.
TITLE The Development of Listening Tests.
PUB DATE 12 Nov 80
NOTE 23p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (New Orleans, LA, November 12, 1980).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Communication Research; *Listening Comprehension; *Listening Comprehension Tests; Listening Skills; *Research Needs; *Test Construction; Test Reliability; Test Reviews; *Test Validity

ABSTRACT

One of the major factors for the lack of quality listening research is a lack of knowledge about the process itself. The second major factor is the lack of a reliable and valid listening test. From a review of 107 articles on listening, 38 reports of listening research, and from studies of seven examiner-developed and three commercially-developed listening tests, a chart was developed to report correlations between each of the listening tests and the intelligence measures and reading assessments used in the validation process or in the research study. The listening tests reviewed include: the Wallner Test of Listening Comprehension, the Tipton and Weaver listening test, the Orr-Graham Listening Test, Sprache's Auditory Comprehension Test, the Brown diagnostic test of listening comprehension, Biggs' diagnostic test of listening effectiveness, Blewett's listening test, Heilman's listening test, the Brown-Carlsen Listening Comprehension Test, the Sequential Tests of Educational Progress (STEP) - Listening, and the Jones-Mohr Listening Test. (Author/PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
C. L. Faires
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED220528

The Development of Listening Tests*

Listening is the most neglected of the communication skills. As large print headlines for a recent full page Wall Street Journal advertisement (Sperry, 1979, p. 171) read, "How can we expect him to learn when we haven't taught him to listen?" Reading, writing, and speaking have long been integral parts of the educational process. They have been included in school curricula, they have been the basis of pedagogical studies, and they have been researched and written about for many years. Each year students, from kindergarten through college, study reading and writing, are taught speech skills and are expected to answer questions, take part in discussions, and learn to express themselves orally. But, while listening is the first language skill to appear chronologically it receives the least amount of attention. The interrelation of each of the lauguage communication skills has been pointed out: "...reading may depend so completely upon listening as to appear to be a special extension of listening" (Lundsteen, 1971, p. 3). Yet Brown (1951) in an early study found 2,600 references on various phases of reading and only 20 references on the subject of listening. This listening area of communication (Brown, 1954; Toussaint, 1960) is neglected as an area of instruction and an area of research (Gold, 1973), and the auditory area is neglected in education (Silverstone, 1968) except in some cases for those with hearing impairments. Brown (1967), in analyzing language arts series, reported that authorities claim that listening is the communication medium most used by children but rarely stressed in texts and even less in the classroom.

One if the major factors for the lack of quality listening research is a lack of knowledge about the process itself. The second major factor is the lack of a

TM 820 550

---

*by Charles L. Faires, Associate Professor, Curriculum and Instruction, The University of Tennessee, Knoxville. Presented at the Mid-South Educational Research Association Annual Convention, New Orleans, Louisiana, November 12, 1980.

reliable and valid listening test. A test has not been regionally nor nationally normed. As early as 1956 it was reported (Biggs, 1956) that, "Although pioneer work has been done in construction of listening tests, the writer was unable to find a study which attempted to validate a diagnostic test against a criterion measurement (p.9)." Shane and Walden (1978) writing twenty years later, stated "A search of the literature as of 1977 failed to unearth a comprehensive diagnostic listening test. Only a few standard tests that provide a standard score are available (p.51)."

Several studies have concerned themselves with the differences between comprehension and a resultant listening test and listening comprehension under more normal conditions. Bakan (1956), suggested, "that training that results in improved scores on a listening test may not necessarily produce improvements in listening under more natural conditions (p.112)." Hacket (1955) in a similar study determined that listening in everyday life was not related to learning to listen for a specific test. A later study by Kelly (1963), again attempted to discriminate the difference between a planned listening test and one given on "surprise" basis or with no notice.

The tests reviewed and reported on in this paper will be of the type where the examinees are aware of the testing conditions, training conditions, or know they will be tested after some cognitive treatment has been performed, immediately or over a period of time.

Several very important problems have surfaced in the literature which have not been considered in listening research and more importantly when developing listening tests.

Research is confusing (Kelly, 1965).

Listening is a complex activity not a unitary skill (Kelly, 1965).

There is a paucity of research in listening (Berger & Werdmann, 1978).

Lundsteen (1969) listed several major difficulties that have hampered listen-

ing research attempts and hinder efforts to develop a truly reliable and valid listening test.

> terms are not clear and communicative
>
> aspects of the teaching of critical listening or its outcomes confound the definition
>
> the ideal of including all behaviors of critical listening in a single definition evades control (p.121).

There are several commercial listening tests available, but as will be pointed out later in this paper, they have not been totally accepted on the basis of their validation procedures and their reliability factors.

This review of research literature will deal with listening tests of two types. First, those tests which are attempts by researchers to validate and prepare reliable listening assessment instruments for all age levels including college will be presented. Secondly, the assessment of commercial listening instruments by those who have used, reviewed, and evaluated the effectiveness of the respective tests will be reported. A series of conclusions concerning the state of the art of listening tests will be included.

In the absence of recognized or accepted listening assessment instruments, those which have been validated or whose reliability has been attested to, several researchers have developed listening tests. Wallner (1974) developed and validated a test of listening comprehension for kindergarten and beginning first graders.

Twelve appropriate passages were written for a specific age level and were paired into two equivalent forms with respect to levels of difficulty (Forms A & B). A panel of appropriate doctoral students judged the level and content validity of the test items. A crayon was used to "x" the correct answer, a line drawing in each case, since children at this age don't read. One hundred fifty students were used as the sample.

Form A of the Test of Listening Comprehension had an internal-consistency reliability of .95 and Form B, .95. The alternate form reliability of the tests

was .89. The criterion-related validity, established by a product-moment correlation with the Metropolitan Readiness Tests was .74 (Form A) and .72 (Form B). Predictive validity, a product-moment correlation with the Standard Achievement Test was .68 and .64, for Form A and Form B respectively.

While not meeting the criterion of statistically equal means, both Form A and Form B did meet the criteria for parallel forms, they had statistically equal variances and intercorrelations.

With regard to the population studied, the Wallner Test of Listening Comprehension appears to have sufficient reliability and validity to warrant continued exploration and use as an effective measure of early listening ability.

In attempting to answer a question, "what are we trying to measure in a listening test?" Tipton and Weaver (1973) developed a listening test for the intermediate elementary grades. They also listed the lack of validity in other listening tests as a primary reason and concern.

Through a comprehensive review of the literature, fourteen listening subskills were chosen as content materials for the analysis. The students listened to the recorded passages and the listening ability was assessed through a multiple-choice test. Sixty-two elementary school students were used as subjects for the analysis. Items were revised after a discrimination index was produced for each item. The test was shortened to 45 minutes and recorded on audio tape. The revised test was administered to 667 students.

This analysis showed 33 of the 35 items to have a significant point bi-serial correlation at the .01 level. These two items were revised and the test again administered to 201 children. All items on the final version adequately discriminated at the .01 level. The test was also compared on the basis of rural and city children. These means were significantly different, (alpha = .05).

The authors reported the internal consistency to be .77 (K-R 20) with no correlation for guessing. The test information also showed an item analysis for

each of the 35 items. Difficulty and discrimination indexes were shown for the fourth, fifth, and sixth grades.

The average difficulty index for each of the three grades was satisfactory, selected sub-skills were deemed appropriate, item analysis showed each item to contribute to the general assessment, the reliability of the test was adequate, and teachers and students reacted favorably to the test and its administrative procedures.

A special listening comprehension test was developed to identify educational potential among disadvantaged junior high school students (Orr & Grahan, 1968). Slower students hold parity with regular students in oral communication up to eleven or twelve years. The reading and vocabulary levels of average and above average students then begins to exceed those of slower students.

Black disadvantaged youths in the Washington, D.C. area were interviewed for their personal interest ideas and the content areas for the test were drawn from interests expressed. Twenty-four content passages were then written and a variety of voice-types were used to record the passages. Twenty junior high school boys were paid for their analysis of the content (introduction, test passages, and closing statement) and for answering the questions. More difficult passages were analyzed by 40 boys. This analysis, based on 90% agreement, resulted in 17 recorded passages and an 86 item listening test. Administration of the final test took ninety minutes.

The listening tests, Forms A and B, were administered to 393 junior high school, eighth grade Black students deemed disadvantaged by family income level. On the three following days the STEP-Reading and the STEP-Listening tests were administered. The Cooperative School and College Ability Test, Level 7-9 (SCAT) was also given to the sample of junior high school students.

The listening test (A&B Forms) had a correlation coefficient of .60 with the SCAT. The total listening test also had a correlation coefficient of .49 with the

STEP Reading and .69 with the STEP Listening Tests. The authors also included an item difficulty index and item-test correlation coefficients for both forms of test and combined totals.

Eighty-one percent of the disadvantaged eighth graders "liked" the listening test and indicated they would prefer it to a reading test covering the same content. Sixty-nine percent indicated that 16 of the 17 passages were interesting and 71% found at least 16 of the passages to be clearly spoken and understandable.

Nunnally (1972) wrote an evaluation for the Mental Measurements Yearbook (1972) of the Orr-Grahan Listening Test and said, "There is no convincing evidence that the test is useful in programs of social improvement, education, career guidance, personnel selection or anything else of an applied nature (p. 636)." He also points out several other problems with the test: (1) the authors did not point out the purpose of the test, (2) the multiple-choice questions mainly required factual recall, and (3) the test required one to write rather than answer oral questions--contradicting an earlier argument for same communication mode stimulus-response. If the test was designed to give Black students an equal advantage, and in the final results the scores were significantly lower than white boys their age, the test has failed.

In the same edition of the Mental Measurements Yearbook, Roger Richards (1972) also evaluates the Orr-Graham Listening Test. The first question he raises is that of validity. The subtitle of the test is "To Identify College Potential Among the Disadvantaged"; he says the test fails in this grandiose idea and therefore is not valid.

The reliability is also questioned. A coefficient of .74 is too low for a published listening test. He also questions the Spearman-Brown coefficient of .85 for the combined forms when nowhere in the literature is there mention of administering the combined forms. A further indicator is a correlation of .69 with the STEP Listening Test; this was also considered too low. Richards also agrees

with Nunnally in pointing out that the introduction of reading into a listening test (the actual test requires reading answers) "...contaminates the measure of listening (p. 637)." In conclusion, Richards said, "The education profession is still without a first-rate instrument for the measurement of listening skill. The Orr-Graham Listening Test has not filled the vacuum."

With only a few attempts to develop listening tests for the elementary and intermediate school levels there seemed to be more interest in test construction and validation at the post high school/college level.

Sprache (1950) believed "...that there was a need for a test to determine the potential ability of students to comprehend and use high-school and college-level reading materials (p. 249)." It was his idea that the listening test should be relatively free from the influence of intelligence. Many intelligence tests are merely an extended reading test or a measurement of reading comprehension. The test was developed in two areas: a Silent Comprehension Test (read and answer questions) and an Auditory Comprehension Test (passages are read to students and then questions are answered).

The results of the Auditory Comprehension Test are correlated with ten tests. The test is also correlated (K-R 21) by total and part scores (on the four areas: Main Ideas, Details, Conclusions, Physical Science) with Diagnostic Reading Tests based on a sample of 162 students.

The reliability coefficient of .79 for the third edition would support the premise that the test would measure those factors operating during reading comprehension.

It was found that a group of above average intelligence high school students could answer about half of the third revision items. The assumption was then made by the author that the remaining twenty-five items were dependent upon the auditorily perceived material in order to properly answer the questions on high school and college textual material.

The author pointed out that the third edition of the test remained parallel by increasing one form to an equal number of items with the others. However, no statistical data on the test, the sample, or item analysis was available in the journal. It seemed the revisions of the first and second editions of both forms guaranteed reliability and validity without further field testing. Without this data and by using a small sample in the beginning its validity and reliability should be questioned.

The purpose of the further development of a listening test was done by Brown (1949) to attempt to meet the basic need for a reliable research instrument. More specifically, the purpose was to develop suitable measuring techniques for use in diagnosing listening comprehension skills among senior high school and college level students.

Based on the assumption that listening is a complex process the test was developed to measure: Section I: Orientation, Oral Directions, Thinking, Strength of Opinion; and Section II: Details, Central Ideas, Depth of Understanding, and Judging Relevancy.

The first edition was tested on 23 students for content validity, ease of following the directions, and total testing time. Some items were dropped and both forms analyzed for uniformity and equality in number of remaining items. The final version was based on the analysis of data collected from 110 high school students and 150 college students.

The analysis to show test validity was done through correlations with existing validated tests. The Brown test, Sections I and II, correlated .37 and .29 with the A.C.E. Psychological Examination, .78 and .67 with the California Test of Mental Maturity, and .31 and .38 with the Nelson-Denny Test of Silent Reading. It was interpreted that Brown's test did not measure intelligence, silent reading comprehension, or scholastic achievement. There was, however, a positive relationship with intelligence. Sections I and II had an intercorrela-

tion of .51.

The reliability was determined by the split-halves method producing a coefficient of .89 and .81 for high school and college levels on Section I and .88 and .65 for Section II.

Interest in the test, appropriateness of the test for use with high school and college students, and clarity of the test purpose was established through interviews with students and teachers.

The author concluded the test to be reliable and valid, useful to measure listening and not intelligence, and listening is a complex activity.

The listening test developed by Biggs (1956) was based on the following rationale, "...in order to understand the listening process, a testing of hypotheses about the factors which may constitute the process is necessary (p. 9)." She felt one method of achieving this was the construction of validated test materials. The purposes of the test are: 1) construct a test which would measure several aspects of college listening, 2) validate the test against outside criteria, and 3) evaluate the test for its independence from factors of reading effectiveness and scholastic aptitude.

The test was administered to 140 college freshmen in a natural classroom setting. Seven tests constituted the battery: 1) grasp central idea, 2) retain pertinent content, 3) recognize main and supporting ideas, 4) hear differences in meaning in similarly-worded statements, 5) recognize correct and incorrect word usage, 6) grasp meaning from contextual clues, and 7) ability to comprehend oral instructions.

The criterion measure used for validation was an instructor rating scale checking for one of three behavior descriptions answering the following questions: 1) Do students grasp oral assignments? 2) How relevant are their questions on oral assignments? 3) Is there an indication of comprehension during discussions? 4) What degree of listening comprehension is determined by answers to questions, and 5) Do students comprehend vocabulary?

Reliability of the test was determined by a rank-difference measure of correlation of ratings (.63); a correlation of two ratings by different instructors (.83); and a correlation coefficient of self-ratings by students over a three week period (.87). The data seemed to have a high degree of reliability. Each of the subtests had reliability coefficients ranging from .75 to .82 except the Content and Inference subtest with a coefficient of .24.

The total test had a correlation coefficient of .52 with the Diagnostic Reading Test and a coefficient of .64 with the A.C.E. Psychological Examination.

The lack of random selection and a fairly subjective rating of students on the criteria scale leave unanswered questions. The criteria scale also called for more than one instructor and self-ratings. However, no inter-rater reliability coefficients are reported. No method for determining validity was mentioned.

Again, in the early 1950's another college level listening test was attempted. Blewett (1951) constructed a listening test designed to measure listening comprehension with validity and reliability and to relate listening skills to aptitudes of 150 college freshmen girls.

The test contained two components: Subtest I: Content Relation and Subtest II: Drawing Conclusions. The Content Relation Subtest contained five passages and took twenty-five minutes to administer. The 40 questions following these passages had a reported mean difficulty index of .46.

The Drawing Conclusions Subtest assessed ability to draw conclusions, make inferences, and identify speaker attitudes. The passages are presented orally and are followed by another 40 item, multiple-choice test with five possible alternatives. The mean difficulty index was .55, interpreted by the writer as easier than Subtest I.

Some simple correlations of the eighty item test (Subtest I and II) and other aptitudinal measures are as follows: A.C.E. Psychological Examination, .51; Psychological Examination Q score, .36; Psychological Examination L score, .73; A.C.E. Reading Comprehension, .39; Purdue Placement Test, .32; and college grade

point average, .37.

The Subtest I - Content Relation and Subtest II - Drawing Conclusions correlated .35 and .52 respectively with the total test and .37 with each other. The only conclusions drawn here by the author (Blewett, 1951) are "...listening on the factual level apparently involves certain factors not operative in inferential listening" and "...the high correlation of .73 between the total test and Psychological Examination indicates the effect of verbal ability on listening concludes the writer (p. 55)."

While this test does not produce any conclusive evidence of the listening process or of the measurement of listening ability the writer suggested, as others have, that the listening process warrants more research.

Another of the college level listening tests developed in the 1950's was one by Heilman (1951). His purpose was to develop a series of training units to improve listening ability. To measure the effect of listening training on subjects compared to those in a control group a test with two comparable forms was developed (Form A and Form B). Six training units were constructed.

The test forms were administered to two randomly selected groups: a control group of 234 college students and an experimental group of 220 college students. Both groups received the Form A as a pretest. The control received no treatment and the experimental group was administered the training units. Then both groups were given Form B of the instrument as a post-test.

The training items were presented via tape recordings once a week for six weeks to the experimental group.

A third subtest, a reading assessment was developed and administered to both groups. The reliability of this test (.84) was determined by the Spearman-Brown Prophecy Formula. The correlation between the results of a reading test and the listening test was .66.

In order to show the relationship of listening to intelligence the American

Council of Education Psychological Examination was administered to the experimental group and a correlation of .56 was calculated.

It was assumed the test measured more than intelligence and that the listening ability of college students can be significantly improved through training. It was also concluded the time spent in the listening training was justified through the results it achieved.

The tests leave a lot to be desired in the way they were constructed and in the reliability and validation results. Test developers have used small samples for determining validity and reliability, they do not report adequate test item analysis data to substantiate their usefulness, and the test studies have not been replicated in other geographical locations or with different samples, (i.e., sex, race, age). A much more rigorous and conservative quantative field testing of these assessment instruments must be made. There are several age groups and school grades, such as lower elementary, for which a test could not be located. Oral examinations must be developed so necessary reading of test items will not contaminate results. These tests and their accompanying materials will have to be made available to interested researchers for further scrutiny.

The aforementioned tests are not normally available and although all are called listening tests by their authors they are not alike, nor do they correlate with each other (Kelly, 1965).

The following three tests: the Brown-Carlsen, the Sequential Test of Educational Progress-Listening, and the Jones-Mohr Test are available on a commercial basis.

The Brown-Carlsen Listening Comprehension Test (1955) has been constructed to measure the ability of students to comprehend spoken language. Listening comprehension is the aural assimilation of spoken symbols in a face-to-face speaker-audience situation, with both aural and visual cues present.

The test has two comparable forms, AM and BM, each with 76 items. Test items

are placed into five groupings, each measuring an important listening skill: (A) Immediate Recall, (B) Following Directions, (C) Recognizing Transitions, (D) Recognizing Word Meanings, and (E) Lecture Comprehension.

The test takes less than an hour, it is administered orally by the proctor, and students record their answers on separate answer sheets which can be scored by hand or machine.

The manual for the exam gives validity and difficulty coefficients for items in each of the five groups for form AM. The test was administered for standardization purposes in 1952, to more than 300 college freshmen. The characteristics of the normative group are based on the Terman-McNear Test of Mental Ability. The intelligence test and the listening test (Form AM) correlated at .70 for the grade 10 norm group and .74 for the grade 11 norm group.

The test was designed to be administered by a teacher without extensive previous training. Clear and precise directions are given in the manual.

The test has been criticized on two points: First, the memory level questions of the trivial numerical type are used almost exclusively on two of the subtests, and Second, the test must be read by the administrator whose individual reading presentation can affect overall outcome. The test, however, does have merit in that in measures the recognition of transitions, i.e., whether a spoken sentence is introductory, transitional, concluding, or none of these. Lindquist (1959) in a critical evaluation of the test reported that no satisfactory evidence of validity has been established for the Brown-Carlsen Test.

One of the more widely used listening tests is the Sequential Tests of Educational Progress (Cooperative Test Division, Educational Testing Service, 1957). Within the Sequential Tests of Educational Progress (STEP) there are six major areas: Reading, Listening, Writing, Social Studies, Science and Mathematics. The four levels of the tests are designed for grades 4-6, 7-9, 10-12, and 13-14 (college level). The designers of the STEP Listening Test use the technique of

having the examiner read the content materials and the multiple choice questions to the examinees. The examinees are given printed copies of the test to read while the questions are being read and they mark their selections on a separate answer sheet.

At each of the four levels the first half of the test consists of thirteen passages which run, timewise, from 25 seconds to just over three minutes. The second half contains exercises more likely to be on a listening test. These are short talks of an argumentative or persuasive nature, oral directions, and instructions and explanations. The manual explains that this material is designed to evaluate several levels of listening cognition: interpretation, identification, remembering, understanding, evaluation, and application.

The proctor or examiner is instructed to devote time to practice in reading the stimulus passages. However, considerable variation can still be expected from this method of presentation of material. Passages are all of the type that could be administered as a silent reading test (Lindquist, 1959, p. 578). Except for the examiner-read passages directions are carefully explained and all materials seem well prepared.

Reviewers of the listening test found no evidence of established validity in the STEP. Reliability coefficients which are reported in the test manual are only those for internal consistency.

Norms provided with the STEP Listening Test are not adequate. One reviewer (Lindquist, 1959) feels a ninety-five percent confidence interval for the 50th percentile should not extend beyond the 45th and the 55th percentiles. However, this is not true in the STEP. The fact that participating schools did not take part in the norming data analysis at all the test levels was also questioned and the schools did not represent a geographic cross section.

Another reviewer of the test (Lorge, 1959) questioned the relationship between the STEP Listening Test and reading and intelligence. This was previously pointed out by Goldstein (1940) who demonstrated the disparity between reading

comprehension and listening comprehension saying both are directly related to a person's intelligence.

The reviewers also thought the test should have longer content passages for subjects to listen to, expecially at the senior high and the college levels. Data to evaluate the difference between listening comprehension and reading comprehension as determined by the STEP was not available.

Kelly (1965) wrote on the construct validity of Brown-Carlsen and the STEP Listening Test. He said, "studies...with a single exception have failed to consider whether the test scores are determined primarily by listening as a unique factor or by one or more general factors, such as intelligence (p. 139)." The same article (Kelly, 1965) also pointed out that both the STEP and the Brown-Carlsen tests had a correlation coefficient of .85 with the OTIS test of mental ability. Both tests purportedly measure listening ability and listening comprehension but only have a correlation coefficient of .82. From this it could be deduced that intelligence accounts for a large portion of listening variance rather than a unique factor of listening that the tests supposedly measure.

The Jones-Mohr Listening Test (1976) developed by John E. Jones and Lawrence Mohr is described as a tape assisted learning program. The Jones-Mohr test is an instrument developed primarily for use in education and training designs and secondarily as an evaluative tool. The test manual lists four purposes for the test.

1. provide immediate feedback to participants on listening accuracy
2. Motivate participants to work on their listening skills
3. demonstrate the need for listening improvement in groups and organizations, and
4. provide a listening instrument for use in research and evaluation studies.

The manual lists the advantages of the test as: brief, easy to administer and score, non-threatening, interesting, and easily adapted to a wide range of education and training situations.

The development of the Jones-Mohr Listening Test was based on listening training and on a set of fifty basic emotions. Each of the emotions were translated into six phrases typical of everyday situations. Male and female actors recorded one sentence from each of the emotions. The most expressive of each of these record phrases was compiled into a list of 300, later edited to 150 statements.

Three forms of the test were assigned fifty items each. The tests were field tested by graduate counseling students. Open ended responses were received on each statement. From these responses, four foils were developed as contrasting multiple-choice answers for each of the items. One foil represented the actual intended meaning and the others were chosen as alternatives. All of the alternatives were somewhat plausible, but one was deemed best. Position of the four alternatives was randomized.

The three forms of the test were revised on the basis of item analysis and sixty items were found to be usable. These items were assigned to matching forms A and B, on a stratified (percent choosing the correct foil) random basis. Further testing resulted in editing on seven additional items.

Norms have not yet been developed for publication. The Jones-Mohr test contains two forms, A and B, with thirty items on each form with four alternatives per item. The instructions are on the tape which is played by the test administrator. Tests are given to each participant, feedback is immediate, and a final score with a range of 0-30 is determined by each person after the test.

The three commercial instruments have many of the same problems as the previously reviewed set of tests. The validation and reliability determining procedures and coefficients are seriously questioned. One major problem is the reading of the content material by the examiner. This introduces a variable that cannot be controlled and norms would therefore be difficult to establish.

All three of the tests need further research and validity and reliability studies. Each is trying to measure listening without the effect of intelligence.

All of the tests except one report a correlation study with a recognized achievement test. None of the authors, however, report using a step-type regression analysis or an analysis of co-variance design to factor out the effects of intelligence.

There is still a controversy between reading and listening. Innovative research is needed to separate, if possible, the elements unique to each and then it may be possible to develop a more sensitive and accurate listening assessment instrument.

The following chart (Fig. 1) shows the reported correlations between each of the listening tests and the intelligence measures and reading assessments used in the validation process or in the research study. The range of coefficients between reading and listening (.31 to .74) closely agree with the range reported by Heilman (1951) .39 to .82.

The need for valid and reliable regionally or nationally normed listening test for all ages and grade levels is exigent. From a review of 107 articles on listening, 38 reports of listening research, and from studies of seven examiner developed and three commercially developed listening tests the following conclusions have been drawn.

there is a great need for more quality listening research

there is a great need for satisfactorily validated and reliable listening tests for all grades and age levels including adults

listening research, for the most part, contains poor research designs which are not always reported nor interpreted correctly in the literature

listening research usually contains non-valid and unreliable instruments for the assessment of listening ability contaminating the research results

many assessments of listening ability are measured with a paper and pencil tests and require some reading ability which also contaminates listening results

more must be known about the listening process itself before adequate assessments can be constructed and validated

two published hierarchies of listening skills (Fessenden, 1955; Weaver & Rutherford, 1974) must be validated and used to define various levels of listening abilities

Fig. 1: Correlations between Listening, Intelligence, and Reading

Listening Test	Reading Tests	Psychological or Intelligence Tests
WALLNER (1974) (Kindergarten and first grade)	Stanford Early School Achievement Test	Metropolitan Readiness Test
Form A	.68	.74
Form B	.64	.52
TIPTON-WEAVER (1973) (Intermediate Elem.)	Not reported	Not reported
STEP: Listening (1957) (Gr. 4-6, 7-9, 10-12, 13-14)	Cooperative Test of Reading Comprehension	Selective Admission Test
	Not reported for all levels or forms	Not reported for all levels or forms
ORR-GRAHAM (1968) (Disadvantaged Jr. High)	STEP: Reading	SCAT
Form A	.52	.60
Form B	.37	.55
A & B	.49	.60
		SCAT Q
Form A		
Form B		.32
A & B		.32
		.31
BROWN (1949) (High school and college)	Nelson-Denny Test of Silent Reading	A.C.E. Psychological Examination
Form A	.31	.37
Form B	.36	.29
BROWN-CARLSEN (1953-55) (High School & College)	Cooperative Test of Comprehensive Reading	Selective Admission Test
	.45	.46
JONES-MOHR (1976) (High school, college and adults)	Not reported	Not reported
SPRACHE (1950) (College)	Diagnostic Reading Test	Terman-McNear I.Q.
	.35	.36
	Cleveland Reading	
	.44	
BLEWETT (1951) (College)	A.C.E. C2-Reading Comprehension	A.C.E. Psychological Examination
	.39	.51
HEILMAN (1951) (College)	Cooperative Test of Reading Comprehension	A.C.E. Psychological Examination
	.66	.56
BIGGS (1956) (College)	Diagnostic Reading Test	A.C.E. Psychological Examination
	.52	.64

listening research studies must be replicated and significant variables must receive further attention

listening variables must be more carefully defined

there is evidence of the significant improvement of listening skills after training sessions designed to improve listening

more attention must be given to developing listening skills and the inclusion of these goals in school curricula

a recent review of media (Faires, 1979) to support listening skill development and instruction showed it to be limited and of poor quality

more readily available and more sensitive item and factor analysis computer programs should be employed on previously developed tests after they are administered to new samples of subjects to reevaluate their worth

Psychologists, educators, and research and development personnel still are in need of a valid and reliable listening test for all grades and age levels. As Brown (1950) said thirty years ago, "...for it is by means such as measures that we may eventually come to a needed understanding of the whole listening process (p. 70)."

REFERENCES

Bakan, P. Some reflections on listening behavior. Journal of Communication, 1956, 6, 111-112.

Berger, A., & Werdmann, A. Listening and auding-activities and research. English Journal, 1978, 67, 36-39.

Biggs, B.P. Construction, validation, and evaluation of a diagnostic test of listening effectiveness. Speech Monographs, 1956, 23, 9-13.

Blewett, T.J. An experiment in the measurement of listening at the college level. The Journal of Communication, 1951, 1, 50-57.

Brown, J.I. Construction of a diagnostic test of listening comprehension. Journal of Exceptional Children, 1949, 18, 139-146.

Brown, J.I. The measurement of listening ability. School and Society, 1950, 71, 69-71.

Brown, J.I. The objective measurement of listening ability. The Journal of Communication, 1951, 1, 44-48.

Brown, J.I. Can listening be taught? College English, 1954, 15, 290-291.

Brown, J.I. & Carlsen, G.R. The Brown-Carlsen Listening Comprehension Test. Yonkers-On-Hudson, New York, World Book Company, 1955.

Brown, K.L. Speech and listening in language arts textbooks: Part 1. Elementary English, 1967, 44, 336-341.

Cooperative Test Division, Educational Testing Services. Sequential Tests of Educational Progress. Princeton, New Jersey: Educational Testing Service, 1957, 1-15.

Faires, C.L. Media for developing listening skills. Unpublished manuscript, The University of Tennessee, Knoxville, 1979.

Fessenden, S.A. Levels of listening: A theory. Education, 1955, 75, 288-291.

Gold, Y. The importance of teaching listening skills. Reading Improvement, 1973, 10, 14-16.

Goldstein, H. Reading and listening comprehension at various controlled rates. Teachers College, Columbia University Contributions to Education, No. 821. New York: Bureau of Publications Teachers College, Columbia University, 1940.

Hackett, H. A null hypothesis: There is not enough evidence. Education, 1955, 75, 350.

Heilman, A.W. An investigation in measuring and improving listening ability of college freshmen. Speech Monographs, 1951, 18, 302-308.

Jones, J.E., Mohr, L. The Jones-Mohr listening test. La Jolla, Calif., University Associates, Inc., 1976.

Kelly, C.M. Mental ability and personality factors in listening. The Quarterly Journal of Speech, 1963, 49, 152-156.

Kelly, C.M. An investigation of the construct validity of two commercially published listening tests. Speech Monographs, 1965, 32, 139-143.

Kelly, C.M. Listening: Complex of activities--and a unitary skill. Speech Monographs, 1967, 34, 455-465.

Lindquist, E.F. Brown-Carlsen listening comprehension test: Evaluation and adjustment series. In O.K. Buros (Ed.), The fifth mental measurements yearbook. Highland Park, New Jersey: Gryphon Press, 1959.

Lindquist, E.F. The sequential test of educational progress: A review. In O.K. Buros (Ed.), The fifth mental measurements yearbook. Highland Park, New Jersey: Gryphon Press, 1959.

Lorge, I. The sequential test of educational progress: A review. In O.K. Buros (Ed.), The fifth mental measurements yearbook. Highland Park, New Jersey: Gryphon Press, 1959.

Lundsteen, S. Critical listening and thinking: a recommended goal for future research. Journal of Research Development in Education, 1969, 3, 119-133.

Lundsteen, S.W. Listening: Its impact on reading and other language arts. Urbana, Illinois: Clearinghouse on the Teaching of English, 1971.

Nunnally, J. Review of the Orr-Graham listening test. In O.K. Buros (Ed.), The seventh mental measurements yearbook. Highland Park, New Jersey: Gryphon Press, 1972.

Orr, D.B., & Graham, W.R. Development of a listening comprehension test to identify educational potential among disadvantaged junior high school students. American Educational Research Journal, 1968, 5, 167-178.

Richards, R.A. Review of the Orr-Graham listening test. In O.K. Buros (Ed.), The seventh mental measurements yearbook. Highland Park, New Jersey: Gryphon Press, 1972.

Shane, H.G., & Walden, V. Classroom relevant research in the language arts. Washington, D.C., Association for Supervision and Curriculum Development, 1978.

Silverstone, D.M. Listening and tape teaching. Audiovisual Instruction, 1968, 13, 870-874.

Sperry Corporation. How can we expect him to learn when we haven't taught him to listen. The Wall Street Journal, November 27, 1979, p. 17.

Sprache, G. The construction and validation of a work-type auditory comprehension test. Educational and Psychological Measurement, 1950, 10, 249-253.

Tipton, M.J., & Weaver, C.H. A listening test for the intermediate grades. Central States Speech Journal, 1973, 24, 5-13.

Toussaint, I. A classified summary of listening. The Journal of Communication, 1960, 10, 125-134.

Wallner, N.K. The development of a listening comprehension test for kindergarten and beginning first graders. Educational and Psychological Measurement, 1974, 34, 391.396.

Weaver, S.W., & Rutherford, W.L. Hierarchy of listening skills. Elementary English, 1974, 51, 1146-50.